DOCUMENT RESUME

ED 173 841 CS 205 005

AUTHOR Peters, Cassandra A.
TITLE Some Communication Problems Related to Testing Minority Populations with Standardized Instruments.
PUB DATE Aug 79
NOTE 26p.; Paper presented at the Annual Meeting of the Association for Education in Journalism (62nd, Houston, Texas, August 5-8, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Communication Problems; Elementary Secondary Education; *Minority Groups; *Standardized Tests; *Test Bias; Test Construction; *Testing Problems; Test Reliability; Test Selection; *Test Validity; Test Wiseness

ABSTRACT

For the most part, standardized tests are verbal and presuppose that the test takers know the language, culture, and values of the test makers, who usually represent white, middle class language and values. The communication problems that minority groups face when taking standardized tests include phonological biases (especially when the tests are orally administered), syntactical biases (unfamiliar sentence structures), semantic/vocabulary biases, cultural biases, and task oriented biases (tasks reflecting a certain set of values). Suggestions for minimizing test bias include validating assumptions made about tests, gathering data concerning the linguistic and cultural backgrounds of the test takers, constructing valid and reliable instruments, training test givers in the use and interpretation of tests, and training test takers. (Appendixes provide validation criteria, suggestions for administering and constructing tests, test taking hints, and what reliability data are essential when reviewing tests. A bibliography is also included.) (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173849

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SOME COMMUNICATION PROBLEMS RELATED TO TESTING MINORITY POPULATIONS WITH STANDARDIZED INSTRUMENTS

By

Cassandra A. Peters
Doctoral Candidate

Department of Communication
Arts and Sciences
Howard University

"PERMISSION TO REPPODUCE THIS MATERIAL HAS BEEN GRANTED BY

Cassandra A. Peters

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

CS205005

Presented to the Minorities and Communication Division of the Association for Education in Journalism Annual Convention in Houston, Texas, August 1979.

Social and behavioral scientists have contributed much to the understanding of human behavior, cognitive functioning, language usage and the like, however, their methodologies have often failed to take into account subcultural differences which govern behavior. Social and behavioral science research and testing methodologies have been largely "ethnocentric" in that they have been derived from a monocultural, normative model which "attempts to assess behavior by using criteria assumed to be universal when they may in fact be idiosyncratic to one group" (Hall and Freedle, 1975, p. 15). Consequently, individuals and groups have been expected to perform in ways that are inconsistent with their past cultural experience. The use of standardized, norm-referenced measuring instruments has led to the misassessment of a wide number of human traits including aptitude, achievement, personality, intelligence, employment potential, and language. Differences in communication patterns among subcultural groups have contributed significantly to this misassessment.

The purpose of this paper is to outline some of the communication problems related to testing minority populations with standardized instruments, to cite relevant literature in support of the claim that tests may be biased against certain groups, and to suggest ways for minimizing or eliminating test bias.

The role of standardized testing has come under serious scrutiny in the past decade. Some of the criticism has erupted due to a generalized decline in test scores across the country

(Thorne, 1976; Rawnsley, 1977). According to Taylor (1977), those who continue to attack standardized tests come from a wide segment of the American population including scholars, parents, teachers, students, legislators, community leaders, media persons, and lawyers who largely represent non-mainstream and minority communities. In a speech delivered before the national conference on "Mobilizing for Excellence in Education," on May 18, 1978, Dr. Mary F. Berry, Assistant Secretary for Education (H.E.W.), made the following comments concerning standardized testing in schools:

> Tests are useful and important and minority children may be taught to pass them as well or better than majority children. But they do not have to be all of what education is about. The day we allow our system of schooling to be bound and defined entirely by objective tests will be the day we abandon creativity of thought as one of our primary goals. It will mean that we have opted for a simple-minded and short sighted accountability process. And it will mean that we have lowered our level of aspiration to the standard of measurable mediocrity. Tests are fine and I support their use but they are only a means of achieving a goal and not the end-all and be-all of education.

These critical comments can be applied to any situation where standardized tests are used routinely to sort out individuals.

Standardized tests represent one type of standardized measuring instrument. They can be defined as published, norm-referenced measures which are administered under standard conditions to a sample of examinees who are supposedly representative of the population for whom the test is intended (Williams, 1975; Perrone, 1977). Traditionally, these tests

have been "standardized by white researchers on white children and do not involve ethnically relevant content" (Williams and Rivers, 1975). The earliest standardized tests were I.Q. tests based on the Army Alpha Tests used to assign soldiers to different tasks. They were used to select-out individuals having certain morals, values, and skills needed for different occupations. The mental testing movement in America grew out of the study of individual differences. In the academic arena, tests were initially used to identify gifted and mentally retarded children.

Historically, American mainstream society has placed much emphasis on testing. Tests have been used broadly to determine readiness for school, to evaluate educational achievement, to determine preparedness for college, to measure potential for graduate study, and to determine aptitude for certain types of jobs. Green (1975, p. 7) summarized the uses of tests:

> Tests are used for selection, for placement, and for diagnosing individual strengths and weaknesses. They are used for prediction. They are used to describe, evaluate, or assess the status of programs for materials or people or organizations. They are used for measuring change or growth.

Our present concern is with the misuses of tests by educators, researchers, and the testing industry. Since the inception of standardized tests, Blacks and other minorities have been labeled as mentally retarded, slow learners, language deficient, culturally deprived and have been placed in special classes or denied jobs because of their performance on certain intelligence, language, and achievement tests. In

a 1975 report, the Commission on Standardized Testing and the Evaluation of Potential Among Minority Group Members (which was appointed by the American Personnel and Guidance Association in 1973), cited research studies and court cases which emphasized the dangers of misclassification. (Williams et al, 1978): Mercer (1972) found 300 percent more Mexican-Americans and 50% more Black Americans to be identified as mentally retarded than would be expected from their proportions in the community. Rein (1974) reported that a company routinely searched its computers for results of tests administered by the Educational Testing Service which the job applicant had taken and added the scores to the applicant's file. However, Guidelines on Employment Selection Procedures issued by the Equal Employment Opportunity Commission on August 1, 1970 prohibits this practice now. In 1973, Slater wrote about two Alabama girls who were ordered sterilized involuntarily by a federally financed program which had determined they were mentally retarded. They were ordered sterilized because "Boys were hanging around." The above incidents were but a few examples of the negative results of depressed scores on standardized tests.

At this point we might ask, what contributes to these low scores? A review of the literature suggests that a primary factor affecting the scores of Blacks, Chicanos, and other minority or non-mainstream groups is test bias. Green (1975, p. 1) stated,

> A biased test is generally understood to be [one] that produces results that are systematically unfair to some group. For this to happen, the test must ordinarily measure variables for that group at least partly distinct from those it measures for other people in the population.

Green cites as an example vocabulary tests which are used as I.Q. tests. Taking a measure of intelligence from a vocabulary test may be suitable for one group but not for another. Test bias concerns validity, i.e., the degree to which a test taps the characteristic it is supposed to measure.

In 1953, Eells (Williams, 1975b, p. 110) suggested three criteria for use in determining whether a test is fair or unfair:

1. It must contain materials common or familiar to all groups to be tested.
2. The language and symbols of the items must be equally familiar to all groups, and
3. The test must be designed to stimulate equal interest and motivation for the various groups.

All types of biases are discussed in the literature: sex bias, e.g., against females; bias against speech disorders, e.g., deafness or cerebral palsy; language bias; and cultural or environmental bias. This paper will focus on language and cultural bias since these constitute the most dominant type of test bias. Taylor (1977) pointed out that in general, most standardized tests are verbal and presuppose that the test taker has a knowledge of the language, culture, and values of the test maker. However, tests are based on a monocultural

model in that they represent essentially the language and values of only one segment of the American population--the white, middle class. The language that is used is standard English which can be defined as that which is taught in schools, represented in writing, and taught to foreigners.

Furthermore, the language of standardized tests is based on a prescriptive, grammatical competence model, i.e., the language of tests is idealized, frozen and artificial and does not represent language as it is spoken in everyday social context. This would pose problems for any test taker regardless of his or her cultural background. Wolfram (1974) hypothesized that the more distant the speaking style of natural conversation is from the language of testing, the more potential linguistic interference there will be for the testee. He says that in terms of this hypothesis we may predict that those items which accentuate the formal standard English may be found to have an effect on all testees confronted with the language style of testing. Wolfram added that since the distance between formal standard English and informal nonstandard English is greater than that between formal and informal standard English, the more serious interference will be caused for the nonstandard speaker.

The remainder of my discussion of language and cultural bias will focus on tests for children because of my current interest in child language. Roberts (1970) explained that in any test a child must comprehend to produce. Comprehension

involves literal comprehension of the test questions and comprehension of the task which is demanded. The testee must comprehend phonological sequences, syntactic structures, lexical items, and sentence meanings. To the extent that the testee's interpretation of the test items and instructions do not match that intended by the test writer and to the extent that his phonological, syntactic, and semantic systems and values are different, this constitutes a language and cultural bias (Roberts, 1970, Rhodes-Hoover et al, 1975).

Studies have shown that children in general have trouble with the wording of tests. Kennedy (1972) observe that many young children attempting to take tests do not understand the general instructions, let alone some of the individual test items. For example, children in the first and second grades taking the California Achievement Test were instructed: Do not turn this page until told to do so. This may be a "simple, straightforward instruction for an adult," however, it can be incomprehensible for a six year old youngster because of at least three linguistic devices which have been shown to be difficult for young children. The you is deleted in until told to do so. Until you are told would be clearer. The passive rather than active voice is used; until I tell you would be better. Finally, a negative is used with the temporal conjunction until instead of the form Turn this page (only) when...

The language patterns or dialects of groups or speech communities vary according to certain sociocultural factors--ethnicity, geographic location, socio-economic status, and situation. So when we talk about language and cultural bias, we may include such groups as geographically isolated Appalacian whites or lower class southern whites who are outside the American mainstream and who speak nonstandard varieties of English as well as Black, Chicano or Chinese Americans.

A test may be phonologically biased in that the examinee's pronunciation system may not match that of the test or the examiner. Rhodes-Hoover, Pollitzer, and Taylor demonstrated phonological bias on various achievement tests. For example, bilingual/bidialectal speakers are to make distinctions among items which are homophones in their speech, i.e., the items are pronounced the same. On the Stanford Early School Achievement Test Level II: Word Reading and the Metropolitan Achievement Test (Primary I, Form F): Word Analysis, students are asked to identify a dictated word from among several words. Had/hat, held/helm, this/these, tap/tab, right/rat, tag/tack, and fur/for are possible homophones. Wolfram, Williams, and Taylor (1974) demonstrated phonological bias on the Wepman Test of Auditory Discrimination which asks speakers to make discriminations between minimal pairs (only one sound is different) such as pin/pen and wreath/reef. Gardner and Gardner (1978) observed that on orally administered reading and achievement tests, a student's dialect may be used against

him when his mastery of skills is evaluated (Gardner, 1978).

Another type of bias is syntactical bias. Rhodes-Hoover et al pointed out that the wording of some of the items or instructions may be highly formal or confusing. For example, the construction "Only the person to whom you make it out" is seen on the Gates-MacGinitie Comprehension Test; the title "Who Has Seen the Wind?-Neither you nor I" appears on the Stanford Early School Achievement Test; instructions such as "None of the following are true except" appear on many multiple-choice tests (Rhodes-Hoover et al, 1975). Wolfram, Williams, and Taylor (1974) noted that children who use nonstandard grammatical forms such as hisself/himself or two cent/two cents have their responses scored as incorrect on the Illinois Test of Psycholinguistic Abilities (ITPA). Lastly, Roberts (1970) demonstrated several levels of syntactic bias on the ITPA, Metropolitan Readiness Tests and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI).

A third type of linguistic bias is semantic or vocabulary bias. Roberts (1970) noted that at least thirteen of the first fifty items on the Peabody Picture Vocabulary Tests (PPVT) are biased against speakers of nonstandard English. Taylor (1978) pointed out that each item on the PPVT carries an approximate age weighting of three months. Thus, if an individual were to fail all thirteen of the potentially biased items, an erroneous 39-month delay would be calculated against the test taker's vocabulary development. A vocabulary test may

ask about items that are unfamiliar to the child (toboggan, vests, armchair, chandelier, anchor, sword, tailor, luxury, manicure, leisure, or familiar under another name (sunglasses vs. shades; train vs. locomotive; wiener vs. hotdog). Vocabulary tests examine for knowledge of verbal labels rather than abstract semantic concepts. Burling (1973) commented that vocabularies of comparable scope seem to be used by individuals of all linguistic communities. "Specialized areas of vocabulary," he added, "may be less well developed in one language than another, and some words may be typically used in one language more than another." The vocabulary differences merely reflect the varied needs and experiences of the people. Anyone's vocabulary reflects his background and interests. People who never talk about something hardly need a word for it. Vocabulary tests measure one kind of verbal experience. Williams (1972) pointed out that a difference in labeling cognitions is to be found in the language of Black children, i.e., many Black children have developed a different or divergent language representation of persons, events, and objects of the world about them and have labeled these phenomena according to what meets approval in their environment. The Gardners (1978) gave an example of an item on a standardized reading test: next to a picture of clothes hanging on the line, the students are asked to circle the correct word: a) pantry b) laundry c) lunch d) lavish.

> Many minority students are not familiar with the word laundry (because) it is not in their operational language system (i.e., sight, listening, and spoken language). They tend to use terms such as dirty clothes, the wash, and soiled clothes. This is a case of differences found in labeling. (Gardner and Gardner, 1978).

Another source of bias is cultural bias which may be present in the test content, the testing situation, and the test format. Wolfram (1974) and Rhodes-Hoover, Pollitzer, and Taylor (1975) described cultural differences related to the testing situation itself. Socio-linguistic factors have considerable influence on the outcome of tests. Standardized tests assume, for example, that the examinee will be willing to provide obvious information ("Show me your nose." or "How many fingers do you have?") to a total stranger--the examiner. These basic social assumptions may violate the interactional rules for some children in some cultures. Linguistically, lower class children may use a more limited, restricted code in the testing situation. (see Bernstein, 1964 for an elaboration of this concept.)

Cazden (1972) observed that pictures are frequently used in tests for children and they often require interpretations by the child. Not all children have the same understanding of pictorial representation, however. Cazden suggested that correct interpretation of test pictures frequently depends on acceptance of particular conventions. She cited an example of an item #47 on the PPVT. The stimulus word is "yawning." The girl on option #2 (correct choice) is shown to be yawning

in the stereotyped way (hand over mouth)--which depicts the middle-class mannerism that has come to be the accepted cartoon for depicting yawning. The girl on option #2 also has her mouth open and certainly might be yawning and is therefore the choice of children who are less aware of convention in pictorial representation.

Taylor (1978) claimed that there is a general cultural bias in terms of the scene and subject matter of the comprehension and vocabulary sections of many tests.

> For example, the comprehension subtest of the Gates-MacGinitie Test has 34 paragraphs that contain few pictures that are working-class oriented. In most of the tests, middle class values are assumed. In the vocabulary subtest of the Stanford Achievement Test, students are asked to respond to the following: If a person does something against the law, he or she is an ambassador, offender, official, or officer. The correct response may not only be the offender. Depending on the respondent's political and social awareness, all the choices may be considered correct. According to the expectations of the test, however, the answer all of the above would be considered incorrect.

Taylor added that on the vocabulary subtest of the Gates-MacGinitie Test, the examinee is asked to pick the best synonyms for inequality. Among the five possible responses--absence, foreign, difference, similarity, and poor--all except absence and similarity may be correct in a cultural group that perceives that difference, poor, and foreign are clearly associated with inequality. Rhodes-Hoover, Pollitzer, and Taylor (1975) singled out item #30 on the Cooperative Primary Tests (Form 12 A) which asks the child to "pick the right answer" in the following selection: Father said, Once there

was a land where boys and girls never grew up. They were always growing. What was father telling?

The truth ____ A lie _____ A story ____.

According to the researchers, the "right" answer could be any of them. Metaphorically, it could be "the truth" if the growth were mental and physical. It could be "a lie" in that the word lie in some varieties of Black speech can also mean a joke or a story; and it could also be "a story."

Roberts (1970), Wolfram (1974) and Cazden (1972) have discussed task bias. Test makers assume that all children have the same presuppositions about the information being requested and that there is only one correct way of responding. This means that the individual is expected to respond within the framework of a certain set of values or norms, e.g., "What would you do if you found a large sum of money on the street?" Middle class values might dictate reporting your discovery to the authorities, but the child might be operating under a different value system. The format of the test is yet another source of potential bias. The test format may block a response to a problem, e.g., if too many distractors (all of the above, none of the above, all are correct, all are incorrect) are used or the response mode is inflexible (answer sheets) or the response options are restricted (Williams et al, 1978; Gardner and Gardner, 1978).

So far we have discussed some of the ill-effects of standardized testing and have identified several potential sources of linguistic and cultural bias. The question remains

as to what to do about these biased instruments. For the remainder of the paper we shall discuss ways in which test bias can be reduced or eliminated. While some researchers have suggested that we throw out tests (Cameron, 1970, Gunning's 1971; Zacharias, 1975; Perrone, 1977; Williams and Mitchell, 1977; Orasnu, 1977), this would not appear to be a viable solution. Culturally fair tests are needed for a variety of reasons: to measure pupil progress; to diagnose learning problems and language disabilities; to evaluate educational/clinical programs and teacher/clinician performance; to support subjective evaluations; to determine the need for additional financial and material resources; to place children in an appropriate educational curriculum; to determine teaching strategies; and, for accountability to parents and school boards. It is therefore suggested that we eliminate the bias and not the tests.

There are several ways in which language and cultural bias can be reduced or eliminated. Taylor (1978) has suggested that once potential sources of bias are identified, that we test our hypotheses in controlled experiments to validate our claim that bias exists. That is, we should determine if a large percentage of examinees fail a particular item. This would indicate bias against a particular group. Once the biased items are determined, they should be eliminated from the test, or the respondent should be given credit for those items, or the biased items should be substituted with fairer

items. These procedures would allow us to make use of tests which are currently in existence, especially if the tests have overall merit, e.g., as valid predictors of achievement or learning potential for mainstream children. Elimination of the unfair items would increase the test's reliability when used with nonmainstream groups. Taylor (1978) further suggested that test makers include items on the new or revised tests to reflect the primary characteristic of the nonmainstream cultures. This, he said, would necessitate the new and revised tests to be standardized on heterogeneous samples of the American population and in some cases it would be necessary to develop a number of norms for a single instrument.

Gardner and Gardner (1978) provided a system to appropriately construct and select tests for Black elementary school pupils which could be used with any subgroup. They stated that efforts to achieve content validity should include review of both the subject matter and kind of behavior or task required from the pupils. The authors outlined some criteria which may be used to validate individual items, clusters of items, or perhaps entire tests which included constructing high interest activities and allowing for a variety of response modes such as discussions, pencil and paper activities, and manipulation of objects. The Gardners concluded their discussion with some suggestions for administering and constructing tests which consisted of exposing disabled readers to remedial instruction; avoiding obscure language; marking on the test booklet itself; designing

power instruments; allowing for flexibility; giving credit for partial answers; and relating test content to the student's learning experiences (See Appendix 1 for an elaboration of the points mentioned).

The criterion-referenced approach to testing has been proposed by many researchers (Region 9 Task Group, 1976; Sweezey, 1975; Williams, 1975; Reichman, 1976). According to Robertson, (1975) a criterion-referenced test is one that is used to identify an individual student's status with respect to an established standard of performance. This standard of performance, a performance objective, describes the expected student changes and the conditions under which these changes will occur. Performance on test items is meant to tell how much the student knows about the topic, how well he or she can perform the task. Criterion-referenced test scores demonstrate mastery of skills whereas norm-referenced tests show the student's position in a sample distribution.

A final major approach to the testing problem is to train the test taker. This solution has been referred to as test wiseness in the literature (Moore, 1971; Oakland, 1971). In an attempt to raise District scores, the Washington, D.C. Public Schools are now beginning a program to train students to take standardized tests (Downey, 1977). Researchers have suggested that training test takers include:

1. familiarizing them with the test situation and the testing format;

2. teaching the language of the test instructions and the test content;
3. developing comprehension and production strategies;
4. expanding their socialization process for increased awareness of mainstream culture and values;
5. teaching task-oriented procedures; and
6. providing hints for being physically, emotionally, and mentally prepared for the testing situation (see Appendix 2 for eight hints).

Conclusion

This paper has outlined some of the directions professionals should take in alleviating the testing problem. In summary, researchers (particularly linguists and other language scientists) should take a more active role in validating the claims and assumptions made about tests and the groups for whom they are intended. They need to gather data concerning the linguistic and cultural backgrounds and values of the testees. Responses considered deviant in the majority culture might be typical of a minority group. Researchers need to review tests for biases and determine their course of action. They should train test makers to construct valid and reliable instruments. Reliability data should be reported for subgroups (see Appendix 2). Competent investigators should train test-givers (educators, psychologists, speech pathologists, counselors) in the use and interpretation of tests. Background factors should be considered in the interpretation of test results. Experts need to devise programs to train the test takers. Finally, local, state, and

national agencies should evaluate their testing programs and revise them where necessary.

APPENDIX 1

VALIDATION CRITERIA
(Gardner and Gardner, 1978)

1. Employ interviews, discussions, and other interactions with either audio-or video-tape records. This would allow the examiner to gain insight into the student's understanding of the content involved and whether the items tap that content.

2. Construct games and other high-interest activities that call for knowledge of a particular content area. Action-oriented or language experience activities and manipulation of concrete objects have proved helpful.

3. Use items with accompanying hand-on manipulative materials that require application of the content learned. This is related to #2 where the student displays skills in working with materials or categorizing them. The examinee could be asked why he grouped certain materials together.

4. Use paper-and-pencil activities other than objective tests. This would give students more latitude in how they go about answering items.

Some Suggestions for Administering and Constructing Tests:

1. Disabled readers should be exposed to reading remediation to assure that they understand content of the test items and the instructions.

2. Test items should be clearly worded avoiding obscure language.

3. Students should be able to mark on the test booklet itself rather than a separate answer sheet.

4. Tests should be designed as power instruments so that most of the pupils will have sufficient time to complete all items.

5. Achievement tests should be designed so that students have some room for flexibility. For example, a cluster of three or four items that examine a particular skill from different viewpoints.

6. The format of test items should be redesigned to give credit for basic or partial understanding of concepts, procedures, processes, and events.

7. Information regarding test structure, technology and uses should be distributed to teachers, parents, and students to facilitate selection of more appropriate tests.

8. Tests should be criterion rather than norm-referenced. Criterion-referenced testing would make it clear that the content sampled by the items relates to the student's learning experiences, and the tests would clearly establish minimum levels of satisfactory performance.

APPENDIX 2

Wolfram (1974) discussed eight hints for taking tests provided by the U.S. Department of Labor in 1968 which would be helpful to anyone taking a standardized test:

1. Get ready for the test by taking other tests on your own.
2. Don't let the thought of taking a test throw you, but being a little nervous won't hurt you.
3. Arrive early, rested, and prepared to take the test.
4. Ask questions until you understand what you are supposed to do.
5. Some parts of the tests may be easier than others. Don't let the hard parts keep you from doing well on the easier parts.
6. Keep time limits in mind when you take a test.
7. Don't be afraid to answer when you are sure you are right, but don't guess wildly.
8. Work as fast as you can, but try not to make mistakes, some tests have short time limits.

Reliability Data
(Fishman, 1964)

1. A test publisher should describe the reliability sample in terms of factors such as age, sex, and grade level composition.
2. The publisher should report subgroup reliabilities.
3. Institutions that use tests regularly on minority groups should make their own reliability studies to determine whether the tests are reliable when used with these groups.

BIBLIOGRAPHY

Bernstein, B. Elaborated and Restricted Codes: Their Social Origins and Some Consequences. American Anthropology, V.66, N.2, 1964, pp. 55-69.

Berry, Mary F. "The Concept of Excellence," a paper delivered at the National Conference on Mobilizing for Excellence in Education, Washington, D.C., May 18, 1978.

Cazden, Courtney. Child Language and Education. New York: Holt, Rinehart, and Winston, 1973, pp. 98-99.

Cameron, Howard K. The Jensen Controversy: II, Cultural Myopia, MEAS Evaluation Guidance, Spring 1970, pp. 10-17.

Downey, Gregg W. Is It Time We Started Teaching Children How to Take Tests. American School Board Journal, V. 164, NI, Jan., 1977, pp. 27-31.

Fishman, Josua et al. Guidelines for Testing Minority Group Children. Journal of Social Issues, V.20, N.2, April, 1964.

Gardner, William E. and Dorothy J. Gardner. Achievement Test Construction and Selection for Black Children. Journal of Non-white Concerns, V.6, N.3, 1978, pp. 148-159.

Goslin, David A. Criticism of Standardized Tests and Testing. New York: College Entrance Examination Board (CEEB), 1967, 34p.

Green, Donald Ross. What Does it Mean to Say a Test is Biased? Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., Mar. 30-Apr. 3, 1975, 25p.

Gunnings, Thomas S. Response to Critics of Robert L. Williams. Counseling Psychologist, V.2, N.3, pp. 73-77.

Hall, William S. and Roy O. Freedle. Culture and Language: The Black American Experience. Washington, D.C.: Hemisphere Publishing Corporation, 1975.

Kennedy, Graeme. The Language of Tests for Young Children. Bernard Spolsky (Ed.) The Language Education of Minority Children, Rowley, Mass: Newbury House Publishers, Inc., 1972, pp. 164-181.

Mercer, J. I.Q.: The Lethal Label. Psychology Today, September, 1972, pp. 44-47, pp. 95-97.

Mitchell, Horace and Robert Williams. A Critical Analysis of the Testing Game. Journal of Non-white Concerns in Personnel and Guidance, V. 6, N.3, April, 1978, pp. 110-118.

Moore, James C. Test Wiseness and Analogy Test Performance. Measurement and Evaluation in Guidance. V.3, N.4, Winter, 1971, pp. 198-202.

Oakland, Thomas and Evelyn Weilaert. The Effects of Test Wiseness Materials on Standardized Test Performance of PreSchool Disadvantaged Children. Paper presented at the Annual Meeting of the American Educational Research Association, New York, Feb., 1971, 13p.

Orasanu, Judith, et. al. A Critique of Test Standardization. Social Policy, V.8, N.2, Sept-Oct, 1977, pp. 61-67.

Perrone, Vito. The Abuses of Standardized Testing. Bloomingdale, Ind.: Phi Delta Kappa Educational Foundation, 1977, 45p.

Region 9 Task Groups on Nonbiased Assessment. A Position Statement on Nonbiased Assessment of Culturally Different Children. Highstown, N.J.: Northeast Area Learning Research Center, Nov., 1976, 42p.

Reichman, Susan L. and Albert C. Oosterhof. Strategy Guidelines for the Construction of Mastery Tests. Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, Ca., April 19-23, 1976), 26p.

Rhodes-Hoover, Mary, Robert L. Pollitzer, and Orlando L. Taylor. Bias in Achievement and Diagnostic Reading Tests: A Linguistically Oriented View. A Paper Presented at the NIE Conference on Test Bias, Dec. 3-5, 1975, Annapolis, Maryland, 28p.

Roberson, Don R. Development and Use of Criterion-Referenced Tests. Austin, Texas: Educational Systems Association, 1975, 28p.

Roberts, Elsa. An Evaluation of Standardized Tests as Tool of Measurement of Language Development. Language Research Reports, No. 1, Cambridge, Mass: Language Research Foundation, 1970.

Slater, J. Sterilization: Newest Threat to the Poor. Ebony, October, 1973, pp. 150-156.

Sweezey, Robert W. and Richard B. Pearlstein. Guidebook for Developing Criterion Referenced Tests. Reston, Va.: Applied Science Associates, 1975, 187p.

Taylor, Orlando. The Sociolinguistic Dimension in Standardized Testing. Georgetown University Roundtable on Languages and Linguistics, 1977, pp 257-266.

Taylor, Orlando. Language Issues and Testing. Journal of Non-white Concerns in Personnel and Guidance, V.6, N.3, April, 1978, pp. 125-133.

Thorne, John M. "Has the Key to the Mystery of Drops in Standardized Test Scores Been Discovered?" Educational Technology, V.16, N.7, July, 1976, pp. 17-18.

Williams, Robert L. Developing Cultural Specific Assessment Devices: an Empirical Rationale. In R.L. Willams (Ed.) Ebonics: The True Language of Black Folks. St. Louis, Mo.: The Institute of Black Studies, 1975, pp. 110-132.

Williams, Robert L. The Problem of Match and Mismatch in Testing Black Children. A paper presented at the Annual Meeting of the American Psychological Association, Honolulu, September, 1972.

Williams, Robert L. and Horace Mitchell. Whatever Happened to ABPsi's Moratorium on Testing: A 1968-1977 Reminder. Journal of Black Psychology, Aug., 1977/Feb., 1978, pp. 24-42.

Williams, Robert L. and L.W. Rivers. The Effects of Language on The Test Performance of Black Children. In R.L. Williams (Ed.) Ebonics: The True Language of Black Folks, St. Louis, Mo.: The Institute of Black Studies, 1975, pp. 96-109.

Williams, Robert et al. An Abridged Version of the 1975 Commission on Standardized Testing Report to APGA. Journal of Non-white Concerns, V.6, N.3, April,1978, pp. 100-109.

Williams, Robert L. et al. Test Format as a Form of Bias for Black Students. Journal of Non-white Concerns, V.6, N.3, April, 1978, pp. 141-147.

Wolfram, Walt. Levels of Sociolinguistic Bias in Testing. Mimeograph dated 1974 to appear in Seminar in Black English, Erbaum Publishers.

Wolfram, W.; Williams, R.; and Taylor, O. Sociolinguistics and Speech Disorders. ASHA Short Course, 1974.

Zacharias, Jerrold R. The Trouble with IQ Tests. National Elementary Principal, V.54, N.4, Mar-Apr. 1975, pp. 23-29.